#3

PATENT
Customer No. 22,852
Attorney Docket No. 03495.0209

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	
Edouard YERAMIAN	Group Art Unit: Not yet assigned
Serial No.: 09/950,051	Examiner: Not yet assigned
Filed: September 21, 2001	
For: GENES AND THE PHYSICS OF THE DNA DOUBLE-HELIX. FORMULATION OF A PHYSICS-BASED GENE IDENTIFICATION (PBGI) METHOD: *AB INITIO* IDENTIFICATION OF GENES IN EUKARYOTIC GENOMES	

Commissioner for Patents and Trademarks
Washington, DC 20231

Sir:

PRELIMINARY AMENDMENT

Prior to the examination of the above application, please amend this application as follows:

IN THE SPECIFICATION:

Please amend the specification by replacing or including the following paragraphs.

Please replace paragraph 11, which begins on page 6 and continues to page 15 of the specification, with the following paragraph:

[011] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

provided by the U.S. Patent and Trademark Office upon request and payment of the necessary fee. This invention will be described in detail with reference to the following drawings:

Fig.1. *Classical model of helix-coil transitions in linear DNA.*

The physics of the transitions is represented schematically, with increasing temperatures. Disruptions in the double-helical structure occur at the extremities of the linear molecule (denaturation from the extremities) as well as within the molecule. Single-stranded loops appear at specific sites, depending on the precise sequence. Because of the cooperativity of the transition, when two loops are close enough they tend to merge together.

Fig. 2. *Helix-coil calculations for the whole yeast chromosome VIII.*

All treatments are for the complete sequence of yeast chromosome VIII, as available in the MIPS databank (http://www.mips.biochem.mpg.de/proj/yeast/).
(A) GC% plot. The calculation is done with a 500 base pairs sliding window.
(B) Stability map with the nearest-neighbour model. The probability for a base pair to be in the coiled state is plotted along the sequence for the temperature T = 63 °C.
(C) Stability map with the long-range effect. The helix-coil calculations are performed with exactly the same parameters, and the same temperature, as in Fig. 2B with the only difference being that it is taken into account of the length-dependent part of the loop-entropy weight. **(D)** Close-up for the detailed stability map with 5 different temperatures. With the same conditions as in Fig. 2C, the stability maps for 5 different temperatures are superposed (63°C-67°C) and a close-up of the corresponding composite map is shown for the chromosonic region extending from 100,000 to 140,000 base pairs. The color coding for the temperatures is displayed with T63, for example, standing for T = 63°C (this convention is adopted throughout the figures).
(E) Superposition of stability and genetic maps. A chromosonic region of 20,000 base pairs long is shown (from 100,000 to 120,000 base pairs), with the positions and orientations of the 8 genes in the region indicated by dark blue arrows. The two superposed stability maps displayed were calculated as in Figs. 2B and 2C, for the

temperature T = 68°C, without (cyan blue) and with (magenta) the loop-entropy long-range effect.

Fig. 3. *Stability maps for Prototheca wickerhamii and Mycobacterium tuberculosis sequences.*

(A) Stability map for the *Prototheca wickerhamii* complete mitochondrial genome. For this genome (Accession number: U02970) the stability map, with 4 different temperatures, is displayed from position 10,000 to 20,000 base pairs. The genes as documented in the annotation are reported, superposed on the stability map, as blue arrows, with the names of the genes. A series of 14 tRNA genes are also reported, coded with the purple color.

(B) Stability map for the *Mycobacterium tuberculosis* complete genome. For this genome (http://www.sanger.ac.uk/Projects/M_tuberculosis/ or http://bioweb.pasteur.fr/GenoList/TubercuList/) the stability map, with 6 different temperatures, is displayed for a 30,000 base pairs long region. For this randomly chosen region, the origin was set arbitrarily (the beginning of the first gene displayed - Rv1331-corresponds to the position 1,500,659 base pairs in the annotation of the genome).

Fig. 4. *Stability properties of a duplicated region in the yeast genome.*

Sequences and annotations are as in the MIPS database.

(A) Stability map, for a duplicated region of yeast (block 9 in chromosome III). The stability map, with 5 different temperatures, is displayed up to the gene YCL035c of this block. **(B)** Stability map, with 'high' temperatures, for the sequence shown in Fig. 4A. The stability map is displayed with 9 different temperatures (4 temperatures in addition to those shown in Fig. 4A). **(C)** Stability map for a portion of the duplicated block 9 in chromosome IV. The stability map is displayed for the same temperatures as in Fig. 4A. Gene YDR516c (green arrow) is homologous to gene YCL040w and YDR518w (red arrow) is homologous to YCL043c. **(D)** Stability map, with 'high' temperatures, for the sequence in Fig. 4C. The 9 temperatures are the same as in Fig. 4B. **(E)** Stability map for the gene YGL235w in chromosome VII. This gene is homologous to YCL040w, and YDR516c.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

Fig. 5. *Stability properties for tandemly duplicated genes in the yeast genome.*

Sequences and annotations are as in the MIPS database.

(A) Stability map for the tandemly duplicated genes YDR038c, YDR039c and YDR040c, in chromosome IV. (B) Stability map for the tandemly duplicated genes YAR027w to YAR033w in chromosome I. The tandemly duplicated genes are represented with pale orange arrows. (C) Stability map, with 'high' temperatures, for the sequence in Fig. 5B.

Fig 6. *Snapshots for stability curves from the chromosome 2.*

For various conditions used for the calculations, and conventions, see the text. The stability curves are plotted for 5 temperatures (56°C to 60°C). The colour coding for the temperatures is displayed with T56, for example, standing for T = 56°C. The plots in black, or red (panel D), correspond to the GC% curves (the y-axis is scaled such as the range [0-1] for the probabilities corresponds to the range [0%-50%] for the GC composition). The GC% curves are calculated with a sliding window of 100 base pairs long (plots in black) or 200 base pairs long (plots in red). Coding regions (as predicted in the database annotation) are represented by red and blue horizontal bars (the alternance of the two colours; is used to clearly distinguish a gene from the previous and next ones). The names of the genes (as in the database annotation of the complete chromosome, Gardner, 1998) are reported above the horizontal bars indicating the putative coding regions.

(A) Stability maps for a sequence from chr2(0_100). **(B)** Stability maps for a sequence from chr2(200 - 300). **(C)** Stability maps for a sequence from chr2(300_400). **(D)** Close-up views for two regions from the stability maps in **(C)**.

Fig. 7. *Analysis of cloned genes.*

Stability maps are represented with the conventions of Fig. 1 (unless otherwise specified).

(A) A gene encoding the mitochondrial phosphate carrier (Bhaduri-McIntosh and Vaidya, 1996; accession: U49381). **(B)** Gamma-glutamylcysteine synthetase gene ('GCS' gene, in blue; Luersen et al., 1999; accession: AJ006966), with nearby another simple gene in the same sequence (in red). **(C)** CTRP gene (Trottein et al., 1995; accession: U34363). **(D)** Pfs230 gene associated with transmission-blocking target

antigen (Williamson et al., 1993; accession: L08135). **(E)** Alpha-tubulin II gene (Holloway et al., 1990; accession: M34390). **(F)** Ca(2+)-ATPase gene (Kimura et al., 1993; accession: X71765). **(G)** A var gene (Reeder et al.; accession: AF134154). **(H)** Pfc2 gene associated with a protein kinase (cdc2-like protein kinase; Ross-Macdonald et al., 1994; accession: X61921). **(I)** Arf gene for ADP-ribosylation factor (Stafford et al., 1996; accession: Z80359). **(J)** A 'SERA' gene (Fox and Bzik, 1994; accession: U08113). **(K)** cpk (kinase) gene (Zhao et al., 1993; accession: X67288). **(L)** Blood stage antigen (41-3) gene (Knapp et al., 1991; accession: M59961). In addition to the standard conditions the stability maps associated with T61 and T62 are drawn as red lines. Alternative splicing has been demonstrated for this gene, yielding three different mRNAs (in addition to the one corresponding to the 9 exons, the following two combinations: 1+2+3+4+8+9 and 1+2+3+4+6+8+9). **(M)** Primase small subunit gene (Prasartkaew et al., 1996; accession: X99254). The cyan lines correspond to the stability maps associated with the temperatures T59.5 to T59.9, by steps of 0.1°C. **(N)** The sequence is the same as in (M), and the two stability maps are associated with the temperature T59.7. The dark purple line corresponds to calculations with interpolations (probabilities evaluated every 20 base pairs) whereas the filled plot in light purple corresponds to calculations without interpolations. **(O)** PfPK4 gene (eIF-2alpha kinase-related enzyme, Mohrle et al., 1997; accession: X94118). A coherent ORF-analysis can be performed with the low-stability region assimilated to an intron (with the original annotation in X94118 replaced by: join(69..388, 698..3440)). **(P)** A gene whose product is thought to be associated to an exported serine/threonine protein kinase (Kun et al. 1997; accession: U40232). **(Q)** Para-aminobenzoic acid synthetase gene (Triglia and Cowman, 1999; accession: AF119554). **(R)** RNA polymerase III largest subunit gene (Li et al., 1991; accession: M73770).

Fig. 8 *Genes from chromosomes 2 and 3 with known similarities.*

Conventions as in Fig. 1 (unless otherwise specified). All exons in green correspond to rectifications or new predictions, suggested by the physics. For rectifications, or new predictions, the coordinates are provided with the same conventions as in the database annotations. For example 'join(n1..n2, n3..n4)'

corresponds to a gene with two exons (n1 to n2, and n3 to n4, respectively), whereas for a gene in the reverse direction the notation 'complement(join(nl..n.2, n3..n4))' is used.

(A) PFC0865w gene (MAL3P7.2, similar to *C. elegans* RNA-binding protein) in chr3(800_900). **(B)** PFC0915w (MAL3P7.12, similar to ATP-dependent RNA helicase) and PFC0920w (MAL3P7.13, similar to *C. elegans* histone H2A variant) genes in chr3(800_900). The very small exon in green (870762..870768) is appended to the second gene (PFC0920w). **(C)** PFB0505c gene (similar to 3-ketoacyl carrier protein synthase III) in chr2(400_500). The rectifications in green replace the exon (460162..460275) by (460162..460206) and the exon (461335..461518) by (461343..461382). **(D)** PFB0425c gene (similar to yeast YMR7 gene) in chr2(300_400). The plot in red corresponds to T61. The annotation in green corresponds to a gene with 6 exons (a to f), with coordinates: join(complement(388524..388560, 388755..388784, 388964..389162, 389448..389618, 389742..390349, 390500..390611)). **(E)** PFC0495w gene (similar to *E. tenella* aspartyl protease) in chr3(400_600) (the gene overlaps the 400_500 and 500_600 stretches, following the conventions here). Calculations are performed (without interpolations) on a sequence extending from positions 498001 bp (taken as the origin for the stability curves in the figure) to 503580 bp, of the chromosome sequence. The magenta lines correspond to 9 temperatures (59.1°C to 60.1°C, by steps of 0.1°C), in addition to the 5 routine temperatures. The rectifications in green concern the exon 1, replaced by exons a, b and c (coordinates (499392..499598, 499711..499755, 499824..499893)), and the exon 2, replaced by exon d (coordinates (499985..500023)). **(F)** Two close-up views of the stability maps in (E).

Fig. 9. *Discovery and annotation of new putative genes in the chromosome 2.* Conventions as in Fig. 1 (unless otherwise specified). Annotations as in Fig. 3.

(A) New putative gene in chr2(100_200), designated as PFB0107c. complement(join(112383..112432, 112612..113075, 113576..113633))

(B) New putative gene in chr2(400_500), designated as PFB0467w. The annotation for this gene is: join(425181..425272, 425733..425855, 425995..426065, 426248..426299,

426531..426543). **(C)** Stability maps for the same sequence as in (B). The probabilities are evaluated every base pair and the origin is set at the first base pair of a 2.7 kb sequence which spans the predicted gene. The plain curve in blue corresponds to the condition T59.2, and all the red lines correspond to the conditions T59.3 to T60, by steps of 0.1°C. **(D)** New putative gene in chr2(600-700), designated as PFB0687c. The annotation for this gene is: complement(join(622777..622840, 622939..622982, 623139..623529, 623715..623944, 624073..624108, 624250..624306). The detailed exon-assembly at the sequence level is displayed in Fig. 6. The outputs for the ORF-analysis and Blast searches (Blastx, with the color keys for the alignment scores corresponding to the NCBI inventions) are displayed below the stability plots. **(E)** New putative gene in chr2(400_500), designated as PFB0503c. The annotation for this gene is: complement(join((457133..457203, 457309..457379, 457461..457589, 457687..457744, 457933..458208, 458447..458585)). The detailed exon-assembly at the sequence level is displayed in Fig. 7. **(F)** New putative genes in chr2(700_800), designated as PFB0827c (exons a to j). The annotation for this gene is: complement(join(728915..728995, 729110..729239, 729359..729448, 729473..729525, 729744..729941, 730413..730544, 731135..731331, 731548..731623, 731818..731921, 732171..732312)).

Fig. 10. *Discovery of new putative genes in the chromosome 3.*
Conventions as in Fig. 1.
(A) New gene in chr3(500_600), designated as PFC0585w. The annotation for this gene is: join(563377..563425, 563508..563537, 564010..564034, 564276..564305, 564434..564520, 564632..564714, 564830..564966, 565077..565174, 565321..565511, 566109..566142, 566316..566337, 566420..566467, 566654..566694, 566786..566921). **(B)** Alignment between the coding sequences of the PFC0585w gene (lower sequence) (SEQ ID NO: 1) and the G408 gene (upper sequence) (SEQ ID NO: 2), see text. **(C)** Alignment between the coding sequences of the PFC0585w gene (lower sequence) (SEQ ID NO: 4) and the G410 gene (upper sequence) (SEQ ID NO: 3), see text. **(D)** Genomic region in chr3(700_800), between the genes PFC0780w (MAL3P6.15, in red, with the gene further extending on the left-side of the figure) and the gene PFC0785c

(MAL3P6.16, in blue). **(E)** Annotation of the exons associated with the stability plots in (D). The 9 exons are appended to PFC0780w, whose new annotation becomes: join(724949..732808, 732909..732980, 733057..733133, 733240..733345, 733503..733549, 733733..333747, 733875..733968, 734067..734116, 734257..734556, 734685..734737). **(F)** New gene in chr3(700_800), designated as PFC0813c, between the genes PFC0810c (MAL3P6.21) and PFC0815c (MAL3P6.22). The annotation for this gene is:
complement(join(758414..759512, 758615..758635, 758952..759027, 759242..759311, 759390..759500, 759840..759907, 760005..760017, 760215..760274, 760475..760525)).

Fig. 11. *Detailed exon-assembly for the gene PFB0687c.*
Sequence-anlysis for the exon-intron structure of the PFB0687c gene, as represented. graphically in Fig. 4D. Exon sequences are represented in blue and intron sequences in green (the rest of the genomic sequence, nor relevant for the analyzed gene, is in black). Start and stop codons (underlined) as well as splice signals are represented in magenta. (SEQ ID NO: 8)

Fig. 12. *Detailed exon-assembly for the gene PFB0503c.*
Sequence-analysis for the exon-intron structure of the PFB0503c gene, as represented graphically in Fig. 4E. Conventions as in Fig. 6. (SEQ ID NO: 9)

Fig. 13. *Low-stability regions within large open reading frames in genes from chromosomes 2 and 3.*
Conventions as in Fig. 1. Annotations as in Fig. 3.
(A) Stability maps associated with the gene PFC0485w. The two exons corresponding to the database annotation are in blue. A new annotation (6 exons, a to f, in green) is also represented. This annotation takes into account of an additional exon at the right-end of the sequence and assimilates the three sharp low-stability regions (within the second exon in blue) to introns. The corresponding new annotation is: join(485498..485941, 486080..487423, 487592..490361, 490491.492454, 492671..493123, 493849..494042). **(B)** Stability maps associated with the gene PFB0530c. The simple gene of the database annotation is represented in blue. A

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

possible rectification for this annotation is represented with three exons in green. The new annotation is: complement(join(477435..477913, 478538..478704, 478991..479079)). **(C)** Stability maps associated with the gene PFB510w. The simple gene as corresponding to the database annotation is in blue. A possible alternative annotation is also represented, with exons in green. A complete gene-assembly as based on this solution is not performed. **(D)** Stability maps associated with the gene PFC0415c. **(E)** Stability maps associated with the gene PFBO540w.

Fig. 14. Experimental confirmation for the physics-based gene predictions (*Plasmodium falciparum*)

The probability of helix opening is calculated along the genomic sequences (chromosome 2 in 14a to 14e, chromosome 3 in 14f for various temperatures (T56 for example standing for the temperature 56°C, the temperatures are relative to standard energetic and thermodynamic parameters for the DNA double-helix (Yerarmian, E. Gene 255,139-150 (2000); Yeramian, E. Gene 255, 151-168 (2000)). The calculations are performed for stretches of 100 kbp (in 14a the origin is set at 600 kbp, in 14b at 800 kbp, in 14c at 400 kpb, in 14d at 600 kbp, in 14e at 500 kbp, and in 14f at 700 kbp). The stable regions are those which remain in the helical state (probability zero to be in the coiled state). The frontiers of the coding regions are shown by vertical arrows. The corresponding uninterrupted genes, or exons, are represented by horizontal bars (in different colors). Detailed annotations for the cloned genes are provided as supplementary information. **(A)** Genes PFB0827c (blue: a PBGI prediction confirmed by sequencing), PFBO830w (red: database annotation) and PFBO833c (green: database annotation for the long exon, the small exon corresponds to a putative missed exon). **(B)** Gene PFB0927c (blue: a PBGI prediction confirmed by sequencing), **(C)** Gene PFB0503c (a PBGI prediction confirmed by sequencing). This prediction is reported as Fig. 4E in Yeramian, E., Gene, 255, 151-168 (2000). When differences are observed between the experimental results (exons in blue) and the predictions (exons 4, 5 and 6), the predicted exons are drawn in green. The same conventions are adopted in Fig. 14f. **(D)** Gene PFB0683w (blue: a PBGI prediction confirmed by sequencing for the 5 first exons). **(E)** Gene PFB0612c (red: original database

annotation, blue: exons predicted by PBGI and confirmed experimentally). **(F)** Gene PFB0780w, with the original annotation corresponding to a simple gene 7973 base pairs long, extending at the left-side of the graph (indicated as a dashed line in red). The 9 exons predicted by PBGI (Fig. 7E in Yeramian, E., Gene, 255, 151-168 (2000)) were confirmed by sequencing (exons in blue, also represented in green for the predictions, whenever differences are observed between predictions and experience).

Fig. 15. Physics-based analysis of the large subunit of RNA polymerase II.

Fig. 16. Analysis of a genomic sequence from *H. sapiens* (Accession No.: AP001754).

Fig. 17. Close-up view of Fig. 16.

Fig. 18. Gene identification for the gene AgProPO of *Anopheles gambiae* (Accession No.: AF031626).

Fig. 19. Physics-based gene analysis of a non-translated gene of *Plasmodium falciparum*.

Fig. 20. Physics-based analysis of the G6PD gene in *Plasmodium falciparum* (Accession No.: X74988).

Fig. 21. Part of the physics-based gene analysis of the *Homo sapiens* gene (Accession AP001754) is presented. In the original genomic sequence the coding region as discovered by the physics-based method are highlighted in blue text (the non-coding regions are in green, and the splice sites in magenta). Bases 287401 to 287941 correspond to SEQ ID NO: 5; bases 288661 to 290581 correspond to SEQ ID NO: 6; and bases 294241 to 295981 correspond to SEQ ID NO: 7.

Please replace paragraph 108, which begins on page 67 and continues to page 69 with the following paragraph:

[108] With the physics-based gene identification scheme, potential new genes are discovered in the is second half, as represented partially (exons in red, in the

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

region 270 to 300 kbp, and further zooming as shown in Fig. 17). Part of the above analysis is presented in Figure 21 in more detail. In the original genomic sequence the coding regions are in green, regions as discovered by the physics-based are highlighted in blue text (the non-coding regions are in green, and the splice sites in magenta).

Please add the Sequence Listing to the specification.

IN THE CLAIMS:

Please amend the claims as follows:

1. (AMENDED) A method for the identification of genes and genetic signals based on the structural properties of a DNA double-helix comprising the following steps:

(A) using the classical physical model of helix-coil transitions;

(B) calculating stability curves, wherein the stability curves are probabilities of opening of the DNA double-helix, along a given sequence[)], by algorithmic methods;

(C) determining the disruption in the linear DNA for different temperatures;

(D) analyzing the stability curves for the detection of genetic signals, wherein the genetic signals are the disruption of the double-helix, or the identification of coding regions that are simple genes, exons in split genes, or regions of high thermal stability; and,

(E) optionally, performing classical sequence analysis based on structural information of donor/acceptor sites, start and codon stops, in correspondence with the frontiers identified in the stability curves and open reading frames analyses for completing the identification of genes.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

2. (AMENDED) The method as claimed in claim 1, wherein the identification is an identification and *ab initio* prediction method of coding regions comprising simple genes , which do not have introns, or of coding regions that comprise exons in split genes, which contain exons, or of coding regions that comprise both simple genes and exons in split genes, in various genomes.

3. (AMENDED) The method as claimed in claim 1 wherein the method is a procedure for the annotation of various genomes that comprise simple genes lacking introns, or that comprise exons in split genes that contain introns, or that comprise simple genes and exons in split genes.

4. (AMENDED) The method as claimed in claim 1 wherein the method is an *ab initio* prediction method for the identification of genetic signals in various genomes that comprise promoters or regulatory sequences that have the propensity to open the DNA double helix, wherein the promoters or regulatory sequences are easily melted regions.

5. (AMENDED) A method for the identification of genes and genetic signals in various genomes, as claimed in claim 1.

6. (AMENDED) The method of claim 5, wherein the genome is an eukaryotic genome.

IN THE DRAWINGS:

Please add new Figure 21, as indicated in the Request for Approval of Drawing Change and the Submission of Formal Drawings.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

REMARKS

Applicant respectfully requests entry of this Preliminary Amendment prior to examination.

The amendment to the specification, removing material in paragraph 108, on pages 68 and 69 is intended to make this material more legible, in response to the requirement in the Notice to File Missing Parts, by including it in a color drawing. Substitute pages are attached which reflect the removal of the sequence information and the amendment to paragraph 108 indicated above. Please note that the removal of the sequence information has left substitute page 68 intentionally blank. No new matter has been added to these substitute pages. Furthermore, new Figure 21, merely adds the sequence information originally found on pages 68 and 69, which is deleted in this Amendment. No new matter is present in Figure 21.

Furthermore, the text added in paragraph 11, page 15, in the Brief Description of the Drawings describing new Figure 21 was copied directly from the material deleted from original paragraph 108. A Request for Approval of Drawing Changes reflecting this addition to the drawings has been filed concurrently. No new matter has been added by this addition.

The addition of the text to the beginning of paragraph 11, on page 6 of the specification, was added to comply with 37 C.F.R. § 1.84(a)(2)(iv), and indicates that the patent application contains color drawings. No new matter has been added by this addition.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

Finally, the addition of the Sequence Listing to the specification and the addition of sequence identifiers to the Brief Description of the Drawings for Figures 10 B and C, 11, 12, and 21 were added to comply with the requirements of 37 C.F.R. §§1.821-1.825. These additions do not add new matter.

The amendments to the claims were made to comply with United States patent practice and do not add new matter.

If there is any fee due in connection with the filing of this Preliminary Amendment, please charge the fee to our Deposit Account No. 06-0916.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.

Dated: February 25, 2002

By: _____

Kenneth J. Meyers
Reg. No. 25,146
Phone: (202) 408-4000
Fax: (202) 408-4400
Email: Ken.Meyers@finnegan.com

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

Appendix to the Preliminary Amendment of February 25, 2002

Please enter the following amendments prior to examination of the application.

IN THE SPECIFICATION

Please replace paragraph 11, which begins on page 6 and continues to page 15 of the specification with the following paragraph:

[011] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the U.S. Patent and Trademark Office upon request and payment of the necessary fee. This invention will be described in detail with reference to the following drawings:

Fig.1. *Classical model of helix-coil transitions in linear DNA.*

The physics of the transitions is represented schematically, with increasing temperatures. Disruptions in the double-helical structure occur at the extremities of the linear molecule (denaturation from the extremities) as well as within the molecule. Single-stranded loops appear at specific sites, depending on the precise sequence. Because of the cooperativity of the transition, when two loops are close enough they tend to merge together.

Fig. 2. *Helix-coil calculations for the whole yeast chromosome VIII.*

All treatments are for the complete sequence of yeast chromosome VIII, as available in the MIPS databank (http://www.mips.biochem.mpg.de/proj/yeast/).

(A) GC% plot. The calculation is done with a 500 base pairs sliding window.

(B) Stability map with the nearest-neighbour model. The probability for a base pair to be in the coiled state is plotted along the sequence for the temperature T = 63 °C.

(C) Stability map with the long-range effect. The helix-coil calculations are performed

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

with exactly the same parameters, and the same temperature, as in Fig. 2B with the only difference being that it is taken into account of the length-dependent part of the loop-entropy weight. **(D)** Close-up for the detailed stability map with 5 different temperatures. With the same conditions as in Fig. 2C, the stability maps for 5 different temperatures are superposed (63°C-67°C) and a close-up of the corresponding composite map is shown for the chromosonic region extending from 100,000 to 140,000 base pairs. The color coding for the temperatures is displayed with T63, for example, standing for T = 63°C (this convention is adopted throughout the figures).

(E) Superposition of stability and genetic maps. A chromosonic region of 20,000 base pairs long is shown (from 100,000 to 120,000 base pairs), with the positions and orientations of the 8 genes in the region indicated by dark blue arrows. The two superposed stability maps displayed were calculated as in Figs. 2B and 2C, for the temperature T = 68°C, without (cyan blue) and with (magenta) the loop-entropy long-range effect.

Fig. 3. *Stability maps for Prototheca wickerhamii and Mycobacterium tuberculosis sequences.*

(A) Stability map for the *Prototheca wickerhamii* complete mitochondrial genome. For this genome (Accession number: U02970) the stability map, with 4 different temperatures, is displayed from position 10,000 to 20,000 base pairs. The genes as documented in the annotation are reported, superposed on the stability map, as blue arrows, with the names of the genes. A series of 14 tRNA genes are also reported, coded with the purple color.

(B) Stability map for the *Mycobacterium tuberculosis* complete genome. For this genome (http://www.sanger.ac.uk/Projects/M_tuberculosis/ or http://bioweb.pasteur.fr/GenoList/TubercuList/) the stability map, with 6 different temperatures, is displayed for a 30,000 base pairs long region. For this randomly chosen region, the origin was set arbitrarily (the beginning of the first gene displayed - Rv1331-corresponds to the position 1,500,659 base pairs in the annotation of the genome).

Fig. 4. *Stability properties of a duplicated region in the yeast genome.*

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

Sequences and annotations are as in the MIPS database.

(A) Stability map, for a duplicated region of yeast (block 9 in chromosome III). The stability map, with 5 different temperatures, is displayed up to the gene YCL035c of this block. **(B)** Stability map, with 'high' temperatures, for the sequence shown in Fig. 4A. The stability map is displayed with 9 different temperatures (4 temperatures in addition to those shown in Fig. 4A). **(C)** Stability map for a portion of the duplicated block 9 in chromosome IV. The stability map is displayed for the same temperatures as in Fig. 4A. Gene YDR516c (green arrow) is homologous to gene YCL040w and YDR518w (red arrow) is homologous to YCL043c. **(D)** Stability map, with 'high' temperatures, for the sequence in Fig. 4C. The 9 temperatures are the same as in Fig. 4B. **(E)** Stability map for the gene YGL235w in chromosome VII. This gene is homologous to YCL040w, and YDR516c.

Fig. 5. *Stability properties for tandemly duplicated genes in the yeast genome.*

Sequences and annotations are as in the MIPS database.

(A) Stability map for the tandemly duplicated genes YDR038c, YDR039c and YDR040c, in chromosome IV. (B) Stability map for the tandemly duplicated genes YAR027w to YAR033w in chromosome I. The tandemly duplicated genes are represented with pale orange arrows. (C) Stability map, with 'high' temperatures, for the sequence in Fig. 5B.

Fig 6. *Snapshots for stability curves from the chromosome 2.*

For various conditions used for the calculations, and conventions, see the text. The stability curves are plotted for 5 temperatures (56°C to 60°C). The colour coding for the temperatures is displayed with T56, for example, standing for T = 56°C. The plots in black, or red (panel D), correspond to the GC% curves (the y-axis is scaled such as the range [0-1] for the probabilities corresponds to the range [0%-50%] for the GC composition). The GC% curves are calculated with a sliding window of 100 base pairs long (plots in black) or 200 base pairs long (plots in red). Coding regions (as predicted in the database annotation) are represented by red and blue horizontal bars (the alternance of the two colours; is used to clearly distinguish a gene from the previous and next ones). The names of the genes (as in the database annotation of the

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

complete chromosome, Gardner, 1998) are reported above the horizontal bars indicating the putative coding regions.

(A) Stability maps for a sequence from chr2(0_100). **(B)** Stability maps for a sequence from chr2(200 - 300). **(C)** Stability maps for a sequence from chr2(300_400). **(D)** Close-up views for two regions from the stability maps in **(C)**.

Fig. 7. *Analysis of cloned genes.*

Stability maps are represented with the conventions of Fig. 1 (unless otherwise specified).

(A) A gene encoding the mitochondrial phosphate carrier (Bhaduri-McIntosh and Vaidya, 1996; accession: U49381). **(B)** Gamma-glutamylcysteine synthetase gene ('GCS' gene, in blue; Luersen et al., 1999; accession: AJ006966), with nearby another simple gene in the same sequence (in red). **(C)** CTRP gene (Trottein et al., 1995; accession: U34363). **(D)** Pfs230 gene associated with transmission-blocking target antigen (Williamson et al., 1993; accession: L08135). **(E)** Alpha-tubulin II gene (Holloway et al., 1990; accession: M34390). **(F)** Ca(2+)-ATPase gene (Kimura et al., 1993; accession: X71765). **(G)** A var gene (Reeder et al.; accession: AF134154). **(H)** Pfc2 gene associated with a protein kinase (cdc2-like protein kinase; Ross-Macdonald et al., 1994; accession: X61921). **(I)** Arf gene for ADP-ribosylation factor (Stafford et al., 1996; accession: Z80359). **(J)** A 'SERA' gene (Fox and Bzik, 1994; accession: U08113). **(K)** cpk (kinase) gene (Zhao et al., 1993; accession: X67288). **(L)** Blood stage antigen (41-3) gene (Knapp et al., 1991; accession: M59961). In addition to the standard conditions the stability maps associated with T61 and T62 are drawn as red lines. Alternative splicing has been demonstrated for this gene, yielding three different mRNAs (in addition to the one corresponding to the 9 exons, the following two combinations: 1+2+3+4+8+9 and 1+2+3+4+6+8+9). **(M)** Primase small subunit gene (Prasartkaew et al., 1996; accession: X99254). The cyan lines correspond to the stability maps associated with the temperatures T59.5 to T59.9, by steps of 0.1°C. **(N)** The sequence is the same as in (M), and the two stability maps are associated with the temperature T59.7. The dark purple line corresponds to calculations with interpolations (probabilities evaluated every 20 base pairs) whereas the filled plot in light purple

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

corresponds to calculations without interpolations. **(O)** PfPK4 gene (eIF-2alpha kinase-related enzyme, Mohrle et al., 1997; accession: X94118). A coherent ORF-analysis can be performed with the low-stability region assimilated to an intron (with the original annotation in X94118 replaced by: join(69..388, 698..3440)). **(P)** A gene whose product is thought to be associated to an exported serine/threonine protein kinase (Kun et al. 1997; accession: U40232). **(Q)** Para-aminobenzoic acid synthetase gene (Triglia and Cowman, 1999; accession: AF119554). **(R)** RNA polymerase III largest subunit gene (Li et al., 1991; accession: M73770).

Fig. 8 *Genes from chromosomes 2 and 3 with known similarities.*

Conventions as in Fig. 1 (unless otherwise specified). All exons in green correspond to rectifications or new predictions, suggested by the physics. For rectifications, or new predictions, the coordinates are provided with the same conventions as in the database annotations. For example 'join(n1..n2, n3..n4)' corresponds to a gene with two exons (n1 to n2, and n3 to n4, respectively), whereas for a gene in the reverse direction the notation 'complement(join(nl..n.2, n3..n4))' is used.

(A) PFC0865w gene (MAL3P7.2, similar to *C. elegans* RNA-binding protein) in chr3(800_900). **(B)** PFC0915w (MAL3P7.12, similar to ATP-dependent RNA helicase) and PFC0920w (MAL3P7.13, similar to *C. elegans* histone H2A variant) genes in chr3(800_900). The very small exon in green (870762..870768) is appended to the second gene (PFC0920w). **(C)** PFB0505c gene (similar to 3-ketoacyl carrier protein synthase III) in chr2(400_500). The rectifications in green replace the exon (460162..460275) by (460162..460206) and the exon (461335..461518) by (461343..461382). **(D)** PFB0425c gene (similar to yeast YMR7 gene) in chr2(300_400). The plot in red corresponds to T61. The annotation in green corresponds to a gene with 6 exons (a to f), with coordinates:
join(complement(388524..388560, 388755..388784, 388964..389162, 389448..389618, 389742..390349, 390500..390611)). **(E)** PFC0495w gene (similar to *E. tenella* aspartyl protease) in chr3(400_600) (the gene overlaps the 400_500 and 500_600 stretches, following the conventions here). Calculations are performed (without interpolations) on

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

a sequence extending from positions 498001 bp (taken as the origin for the stability curves in the figure) to 503580 bp, of the chromosome sequence. The magenta lines correspond to 9 temperatures (59.1°C to 60.1°C, by steps of 0.1°C), in addition to the 5 routine temperatures. The rectifications in green concern the exon 1, replaced by exons a, b and c (coordinates (499392..499598, 499711..499755, 499824..499893)), and the exon 2, replaced by exon d (coordinates (499985..500023)). **(F)** Two close-up views of the stability maps in (E).

Fig. 9. *Discovery and annotation of new putative genes in the chromosome 2.* Conventions as in Fig. 1 (unless otherwise specified). Annotations as in Fig. 3. **(A)** New putative gene in chr2(100_200), designated as PFB0107c. complement(join(112383..112432, 112612..113075, 113576..113633)) **(B)** New putative gene in chr2(400_500), designated as PFB0467w. The annotation for this gene is: join(425181..425272, 425733..425855, 425995..426065, 426248..426299, 426531..426543). **(C)** Stability maps for the same sequence as in (B). The probabilities are evaluated every base pair and the origin is set at the first base pair of a 2.7 kb sequence which spans the predicted gene. The plain curve in blue corresponds to the condition T59.2, and all the red lines correspond to the conditions T59.3 to T60, by steps of 0.1°C. **(D)** New putative gene in chr2(600-700), designated as PFB0687c. The annotation for this gene is: complement(join(622777..622840, 622939..622982, 623139..623529, 623715..623944, 624073..624108, 624250..624306). The detailed exon-assembly at the sequence level is displayed in Fig. 6. The outputs for the ORF-analysis and Blast searches (Blastx, with the color keys for the alignment scores corresponding to the NCBI inventions) are displayed below the stability plots. **(E)** New putative gene in chr2(400_500), designated as PFB0503c. The annotation for this gene is: complement(join((457133..457203, 457309..457379, 457461..457589, 457687..457744, 457933..458208, 458447..458585)). The detailed exon-assembly at the sequence level is displayed in Fig. 7. **(F)** New putative genes in chr2(700_800), designated as PFB0827c (exons a to j). The annotation for this gene is:

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

complement(join(728915..728995, 729110..729239, 729359..729448, 729473..729525, 729744..729941, 730413..730544, 731135..731331, 731548..731623, 731818..731921, 732171..732312)).

Fig. 10. *Discovery of new putative genes in the chromosome 3.*

Conventions as in Fig. 1.

(A) New gene in chr3(500_600), designated as PFC0585w. The annotation for this gene is: join(563377..563425, 563508..563537, 564010..564034, 564276..564305, 564434..564520, 564632..564714, 564830..564966, 565077..565174, 565321..565511, 566109..566142, 566316..566337, 566420..566467, 566654..566694, 566786..566921). **(B)** Alignment between the coding sequences of the PFC0585w gene (lower sequence) (SEQ ID NO: 1) and the G408 gene (upper sequence) (SEQ ID NO: 2), see text. **(C)** Alignment between the coding sequences of the PFC0585w gene (lower sequence) (SEQ ID NO: 4) and the G410 gene (upper sequence) (SEQ ID NO: 3), see text. **(D)** Genomic region in chr3(700_800), between the genes PFC0780w (MAL3P6.15, in red, with the gene further extending on the left-side of the figure) and the gene PFC0785c (MAL3P6.16, in blue). **(E)** Annotation of the exons associated with the stability plots in (D). The 9 exons are appended to PFC0780w, whose new annotation becomes: join(724949..732808, 732909..732980, 733057..733133, 733240..733345, 733503..733549, 733733..333747, 733875..733968, 734067..734116, 734257..734556, 734685..734737). **(F)** New gene in chr3(700_800), designated as PFC0813c, between the genes PFC0810c (MAL3P6.21) and PFC0815c (MAL3P6.22). The annotation for this gene is:

complement(join(758414..759512, 758615..758635, 758952..759027, 759242..759311, 759390..759500, 759840..759907, 760005..760017, 760215..760274, 760475..760525)).

Fig. 11. *Detailed exon-assembly for the gene PFB0687c.*

Sequence-anlysis for the exon-intron structure of the PFB0687c gene, as represented. graphically in Fig. 4D. Exon sequences are represented in blue and intron sequences in green (the rest of the genomic sequence, nor relevant for the analyzed gene, is in

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

black). Start and stop codons (underlined) as well as splice signals are represented in magenta. (SEQ ID NO: 8)

Fig. 12. *Detailed exon-assembly for the gene PFB0503c.*

Sequence-analysis for the exon-intron structure of the PFB0503c gene, as represented graphically in Fig. 4E. Conventions as in Fig. 6. (SEQ ID NO: 9)

Fig. 13. *Low-stability regions within large open reading frames in genes from chromosomes 2 and 3.*

Conventions as in Fig. 1. Annotations as in Fig. 3.

(A) Stability maps associated with the gene PFC0485w. The two exons corresponding to the database annotation are in blue. A new annotation (6 exons, a to f, in green) is also represented. This annotation takes into account of an additional exon at the right-end of the sequence and assimilates the three sharp low-stability regions (within the second exon in blue) to introns. The corresponding new annotation is: join(485498..485941, 486080..487423, 487592..490361, 490491.492454, 492671..493123, 493849..494042). **(B)** Stability maps associated with the gene PFB0530c. The simple gene of the database annotation is represented in blue. A possible rectification for this annotation is represented with three exons in green. The new annotation is: complement(join(477435..477913, 478538..478704, 478991..479079)). **(C)** Stability maps associated with the gene PFB510w. The simple gene as corresponding to the database annotation is in blue. A possible alternative annotation is also represented, with exons in green. A complete gene-assembly as based on this solution is not performed. **(D)** Stability maps associated with the gene PFC0415c. **(E)** Stability maps associated with the gene PFBO540w.

Fig. 14. Experimental confirmation for the physics-based gene predictions (*Plasmodium falciparum*)

The probability of helix opening is calculated along the genomic sequences (chromosome 2 in 14a to 14e, chromosome 3 in 14f for various temperatures (T56 for example standing for the temperature 56°C, the temperatures are relative to standard energetic and thermodynamic parameters for the DNA double-helix (Yerarmian, E. Gene 255,139-150 (2000); Yeramian, E. Gene 255, 151-168 (2000)). The calculations

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

are performed for stretches of 100 kbp (in 14a the origin is set at 600 kbp, in 14b at 800 kbp, in 14c at 400 kpb, in 14d at 600 kbp, in 14e at 500 kbp, and in 14f at 700 kbp). The stable regions are those which remain in the helical state (probability zero to be in the coiled state). The frontiers of the coding regions are shown by vertical arrows. The corresponding uninterrupted genes, or exons, are represented by horizontal bars (in different colors). Detailed annotations for the cloned genes are provided as supplementary information. **(A)** Genes PFB0827c (blue: a PBGI prediction confirmed by sequencing), PFBO830w (red: database annotation) and PFBO833c (green: database annotation for the long exon, the small exon corresponds to a putative missed exon). **(B)** Gene PFB0927c (blue: a PBGI prediction confirmed by sequencing), **(C)** Gene PFB0503c (a PBGI prediction confirmed by sequencing). This prediction is reported as Fig. 4E in Yeramian, E., Gene, 255, 151-168 (2000). When differences are observed between the experimental results (exons in blue) and the predictions (exons 4, 5 and 6), the predicted exons are drawn in green. The same conventions are adopted in Fig. 14f. **(D)** Gene PFB0683w (blue: a PBGI prediction confirmed by sequencing for the 5 first exons). **(E)** Gene PFB0612c (red: original database annotation, blue: exons predicted by PBGI and confirmed experimentally). **(F)** Gene PFB0780w, with the original annotation corresponding to a simple gene 7973 base pairs long, extending at the left-side of the graph (indicated as a dashed line in red). The 9 exons predicted by PBGI (Fig. 7E in Yeramian, E., Gene, 255, 151-168 (2000)) were confirmed by sequencing (exons in blue, also represented in green for the predictions, whenever differences are observed between predictions and experience).

Fig. 15. Physics-based analysis of the large subunit of RNA polymerase II.

Fig. 16. Analysis of a genomic sequence from *H. sapiens* (Accession No.: AP001754).

Fig. 17. Close-up view of Fig. 16.

Fig. 18. Gene identification for the gene AgProPO of *Anopheles gambiae* (Accession No.: AF031626).

Fig. 19. Physics-based gene analysis of a non-translated gene of *Plasmodium falciparum*.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

Fig. 20. Physics-based analysis of the G6PD gene in *Plasmodium falciparum* (Accession No.: X74988).

Fig. 21. Part of the physics-based gene analysis of the *Homo sapiens* gene (Accession AP001754) is presented. In the original genomic sequence the coding region as discovered by the physics-based method are highlighted in blue text (the non-coding regions are in green, and the splice sites in magenta). Bases 287401 to 287941 correspond to SEQ ID NO: 5; bases 288661 to 290581 correspond to SEQ ID NO: 6; and bases 294241 to 295981 correspond to SEQ ID NO: 7.

Please replace paragraph 108, which begins on page 67 and continues to page 69, with the following paragraph:

[108] With the physics-based gene identification scheme, potential new genes are discovered in the is second half, as represented partially (exons in red, in the region 270 to 300 kbp, and further zooming as shown in Fig. 17). Part of the above analysis is presented in [below] in Figure 21 in more detail. [In the original genomic sequence the coding regions are in green, regions as discovered by the physics-based are highlighted in blue text (the non-coding regions are in green, and the splice sites in magenta):

```
287401 ataaacattc tttagtccac acatagataa ataaataagg aagcaaatag acacacagaa
287461 gagcgggaca gctcctcctc ccgggagaat ttcaattagt aagtgtggaa ggaacaaggc
287521 agggaggaga atcctcaaca gagccccaca gggaccgtgc gggcgaggcc cccggagggg
287581 caccagcact gccgggcaaa cgcctgggca gacgcgggac agctgccaag tctcagacat
287641 gaccaattac agagggaaac ggcggcaccg cgagggatgg gccgcggccg tgtcacctcc
287701 atgccccacg cacactgctc ctgtgggatt cctcccccaa cacgatgccc actctgacca
287761 cgaggaaacc tcaagcaagt ccacgtggag gggcattcta caaaacaccc aaccggtcaa
287821 ggtcgctgag gccaaggaga gattgggcaa ccgtcacaaa ccagagaagc cgaggagagc
287881 tttcagccaa cgccatgtgg ggtcctgagc aggacccacc ggaagttggt gcagctgcct
287941 aaagaccgtc ctggctgaga agaaacagag cagcgctgct ttctcagagc tgggaaccga


288661 acctcgatct cagacttctg gcttccaaaa ccatgagaca cggaatttct gttgtgtgac
288721 cagccagttt gtggtactgt ttgtcatggc agcccaagga aaagaataca ttacagcata
288781 caaaccatga ctcacattat ctttacttag aacccaaaca aacctctctc cctaagcttt
288841 caatcacaga ggcacatgat cttgttcagc agcctagaaa accaaggccc agcggagcca
288901 cccgtaggca cccactcccc atagcctggc acacacacac ggcagagcca cccacaggca
```


FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

```
288961 cccactcctc atagtccagc acacacacgg cagagccacc cgcaggcacc cactccccat
289021 agcccggcac acacgtggac catgccaccc tccacgtgcg cctggggagc aaagcagcac
289081 agcctgaact gcccctcagc tcttcctcct gagtctaaaa cacgcacatg cgccccaggc
289141 caattccaag ttttgtaaac tgagcaacag ctcttgggaa acaaaaacac agctactgtt
289201 tattctcctg gagctggctg tacaccccaa caaggaaggg agggcttgct gagcctcctg
289261 tctggacaac atgcaccaag gaggagtata aaagccccac aaacccgagc acctcactca
289321 ctcgctcacc cactccctcc catctccccc agctcaaccc ccagcacagc agcatccacc
289381 atgtccgtct gctccagcga cctgagctac agcagccgcg tctgccttcc tggttcctgt
289441 gactcttgct ccgactcctg gcaggtggac gactgcccag agagctgctg cgagcccccc
289501 tgctgcgccc ccagctgctg cgccccggcc ccctgcctga gcctggtctg caccccagtg
289561 agccgtgtgt ccagcccctg ctgcccagtg acctgcgagc ccagcccctg ccaatcaggc
289621 tgcaccagct cctgcacgcc ctcgtgctgc cagcagtcta gctgccagct ggcttgctgt
289681 gcctcctccc cctgccagca ggcctgctgc gtgcccgtct gctgcaagac tgtctgctgc
289741 aagcctgtgt gctgtgtgcc cgtctgctgt ggggattctt catgctgcca gcagtctagc
289801 tgccagtcag cttgctgcac ctcctccccc tgccagcagg cctgctgtgt gcccatctgc
289861 tgcaagcctg tctgctctgg gatttcctct tcgtgctgcc agcagtctag ctgtgtgagc
289921 tgtgtgtcca gcccctgctg ccaggcggtc tgtgagccca gcccctgcca atcaggctgc
289981 atcagctcct gcacgccctc gtgctgccag cagtctagct gccagccggc ttgctgcacc
290041 tcctcctcct gccagcaggc ctgctgcgtg cccgtctgct gcaagactgt ctgctgcaag
290101 cctgtgtgct ctgaggattc ctcttcatgc tgccagcagt ctagctgcca gccggcttgc
290161 tgcacctcct ctccctgcca gcaggcttgc tgtgtgcctg tctgctgcaa gcctgtgtgc
290221 tgcaagcctg tcggctctgt gcccatctgc tctggggctt cctctctgtg ctgccagcag
290281 tctagctgcc agccagcttg ctgcacctcc tcccaaagcc agcagggctg ctgcgtgccc
290341 gtctgctgca agcctgtgag ctgtgtgcct gtttgctctg gggcttcctc ttcatgctgc
290401 cagcaatcta gctgccagcc agcttgctgc accacctcct gctgcagacc ctcctcctcc
290461 gtgtccctcc tctgccgccc cgtgtgcagg cccgcctgct gcgtgcccgt cccttcctgc
290521 tgtgctccca cctcctcctg ccaacccagc tgctgccgcc cagcctcctg cgtgtccctc
290581 ctctgacgcc ccgtgtgctc ccgcccagcc tgctgaggcc tccgctcagg tcagaagccc

294241 ggatgagagg gggactcatg gaggaacagc cacgccttga ccctgagatg gccttgcagg
294301 gagggtaact gaaaatttac ccactgggga cagttgccta cttactaaaa cagttccagc
294361 caccaccgca gcccctggaa ggccatcccc ccagaaaatc ccccaggtct cagcagggcc
294421 ttgtccacct gtgccctcca gtgtcgccca tgtcaacctc acctaagagg ggcctgacgc
294481 acggtcctgc aggtgcggac tctgggtcct gacagcccat gcggaacctg gtgcccccag
294541 aggagggcct ggggcagtgc cagttttggg gaatcatgtg catccatcca cccactccat
294601 gatgctttcg tcctgatcga gtcccttgtc tcccgcgcag gtgcagcagc ccctccctct
294661 ccccccgcat tgctgctgaa cgggcagaac cctcgggcgg gcggcacaca gggagggtga
294721 ccaggcctgg aggctgtagt gcccggaccc caggccagct tcctggaagg tgaccctgca
294781 gggtgggctc tcccaggtgg gacagtgggt gggacagtcc tggggcctgg agagccccac
294841 agcccagggc acggcagcca atgaccaggc tcaggaagac ccaggcatgg aggctgagcc
294901 gggactgagc cttcctgggc gtggctgtga gttccacctg gtgacccccc ggaggagtta
294961 ggccactgtc ccccgtgact tctaggttaa gtcactcatt catagaaaca gtcatggcta
295021 gagagcaatc tgagctcaaa accatgtatc cccaggagca ctacagaaaa agagaatcag
295081 gcgaccaagg ggagtttatt ggggagcagg aggaggtgct gacaggttca agtcgaggcc
295141 aagtgacctg gggcagagaa gctgggaggg aggacagggg acccaacagg caggtgggcc
295201 cctgctggga ggcaggagct ggggagcttc gaggatggag attcctggga gtatggaggg
295261 gggggtcacc tcagcacatg ggggccccgt cccaagcggg ggcaacctcc taacccgagt
295321 caggaccagt tggccctggg ggatgtgcac atcagcaact ggactcctgg cctgagcaga
295381 ggcctcagca ggccaggcgg gagcacgcgg ggcggcagag gagggacacg caggaggccg
295441 ggcggcagca gctggcctgg taggaggagg caggggcaca gcaggaggag atgggcacgc
295501 agcaggcggg cctgcatatg gggcggcaga ggagggacac ggaggaggag ggtctgcagc
295561 aggaggtggt gcagcaagcc ggctgacagc tagactgctg gcagcatgaa gtggaagccc
295621 cagagcagac gggcacacag cagatgggtt tgaagcagac aggcttgcaa cagacaggca
295681 cgtagcagga ctgctggcag ggggaggagg tgcagcaagt cggctggcag ctagaatgct
295741 ggcagcatga agaggaatcc tcagaacagg tgggcacaca gcacacgggc ttgcagcaga
295801 caggcacaca gcaggactgc tggcaggagg aagaggcaca gcaagttggc tggcagctag
295861 actgctggca gcatgaagag gaatccttag agcaggtggg caggcagcac acaggcttgc
295921 agcagacggg cacgcagcag gcctgctggc agggggagga ggcgcagcaa gccggctggc

295981 agcacgaggg cgtgcaggag ctggtgcagc ctgattggca ggggctgggc tcacaggccg]
```


IN THE CLAIMS:

Please amend the claims as follows:

1. (AMENDED) A [Method] method for the identification of genes and genetic signals based on the structural properties of a DNA double-helix comprising the following steps:

(A) [-] using the classical physical model of helix-coil transitions[,];

(B) [-] calculating stability curves [(], wherein the stability curves are probabilities of opening of the DNA double-helix, along a given sequence[)], by algorithmic methods[,];

(C) [-] determining the disruption in the linear DNA for different temperatures[,];

(D) [- analysing] analyzing the stability curves for the detection of genetic signals [(], wherein the genetic signals are the disruption of the double-helix[)], or the identification of coding regions [(] that are simple genes, [or] exons in split genes, or regions of high thermal stability[), and optionally]; and,

(E) [-based on the structural informations,] optionally, performing classical sequence analysis [(] based on structural information of donor/acceptor sites, start and codon stops, in correspondence with the frontiers identified in the stability curves and open reading frames analyses[)] for completing the identification of genes.

2. (AMENDED) The [Method] method [according to] as claimed in claim 1, [characterized by an] wherein the identification is an identification and [ab initio] *ab initio* prediction method of coding regions comprising simple genes [(] , which do not have [without] introns[)], [and/or] or of coding regions that comprise exons in split genes [(containing], which contain exons[)], or of coding regions that comprise both simple genes and exons in split genes, in various genomes.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com

3. (AMENDED) [Method according to] The method as claimed in claim 1 [and claim 2 characterized by] wherein the method is a procedure for the annotation of [the] various genomes that [comprising] comprise simple genes [(without] lacking introns[) and/or], or that comprise exons in split genes [(containing] that contain introns,[)] or that comprise simple genes and exons in split genes [in various genomes characterized by performing the steps A to E of claim 1].

4. (AMENDED) [Method according to] The method as claimed in claim 1 [characterized by] wherein the method is an [ab initio] *ab initio* prediction method for the identification of genetic signals in various genomes that [comprising] comprise promoters or regulatory sequences [characterized by] that have the propensity [of] to [opening] open the DNA double helix [which are easily melt-region], wherein the promoters or regulatory sequences are easily melted regions.

5. (AMENDED) [Use of the method as claimed in claims 1 to 4] A method for the identification of genes and genetic signals in various genomes, as claimed in claim 1.

6. (AMENDED) [Use of the method according to claim 5] The method of claim 5, wherein the genome is an eukaryotic genome.

FINNEGAN
HENDERSON
FARABOW
GARRETT &
DUNNER LLP

1300 I Street, NW
Washington, DC 20005
202.408.4000
Fax 202.408.4400
www.finnegan.com